

1. Big Data con AWS

1.1. Introducción a AWS

Amazon Web Services (AWS) es una plataforma de servicios en la nube líder en el mundo que ofrece una amplia gama de servicios de infraestructura, almacenamiento, cómputo, bases de datos, análisis y más.

AWS proporciona la infraestructura necesaria para ejecutar aplicaciones, almacenar datos y procesar información a escala global. En el contexto de Big Data, AWS ofrece herramientas y servicios para administrar, procesar y analizar grandes volúmenes de datos de manera eficiente y escalable.

AWS ofrece una amplia gama de servicios y herramientas para facilitar proyectos de Big Data. Puedes aprovechar la escalabilidad, la variedad de servicios y las medidas de seguridad de AWS para administrar y analizar grandes volúmenes de datos de manera efectiva.

1.2. Por Qué AWS para Big Data

1.2.1. Escalabilidad

AWS ofrece la capacidad de escalar recursos de manera flexible según las necesidades de tu proyecto de Big Data. Puedes aumentar o reducir la capacidad de cómputo y almacenamiento según la demanda, lo que te permite manejar grandes volúmenes de datos sin preocuparte por la infraestructura subyacente.

1.2.2. Variedad de Servicios

AWS proporciona una amplia gama de servicios específicamente diseñados para el procesamiento y análisis de Big Data. Estos servicios incluyen Amazon EMR (Elastic MapReduce), Amazon Redshift, AWS Glue, Amazon Kinesis y más.

1.2.3. Seguridad y Cumplimiento

AWS implementa medidas de seguridad y cumplimiento rigurosas para proteger tus datos. Puedes administrar el acceso a tus recursos y datos, cifrar la información en tránsito y en reposo, y cumplir con regulaciones específicas de la industria.

1.3. Principales Servicios de AWS para Big Data

1.3.1. Amazon S3 (Amazon Simple Storage Service)

Amazon S3 es un servicio de almacenamiento en la nube altamente escalable, seguro y confiable que permite a las empresas almacenar, gestionar y acceder a datos de manera eficiente en la plataforma de AWS.

1.3.2. Amazon EMR (Elastic MapReduce)

Amazon EMR es un servicio de administración de clústeres de código abierto que te permite procesar grandes volúmenes de datos de manera distribuida utilizando frameworks como Apache Hadoop, Apache Spark y Apache Hive. Puedes crear y administrar clústeres de EMR de manera sencilla y escalarlos según sea necesario.

1.3.3. Amazon Redshift

Amazon Redshift es un servicio de data warehousing completamente administrado que te permite analizar grandes conjuntos de datos de manera eficiente. Es ideal para almacenar y consultar datos analíticos, y admite integración con herramientas de visualización populares.

1.3.4. Amazon Athena

Amazon Athena es un servicio de análisis de datos en AWS que permite realizar consultas SQL directamente sobre datos almacenados en Amazon S3, sin necesidad de infraestructura. Es sin servidor y cobra solo por las consultas realizadas. Soporta múltiples formatos de datos y se integra fácilmente con otros servicios de AWS, facilitando el análisis y la visualización de información.

1.3.5. AWS Glue

AWS Glue es un servicio de ETL (Extract, Transform, Load) completamente administrado que facilita la preparación y transformación de datos para análisis. Puedes definir trabajos de ETL utilizando un entorno visual o escribir scripts en lenguaje Python.

1.3.6. Amazon Kinesis

Amazon Kinesis es una plataforma para el streaming de datos en tiempo real. Te permite capturar, procesar y analizar flujos de datos en tiempo real, lo que es fundamental para aplicaciones de análisis en tiempo real y procesamiento de eventos.

1.4. Ejemplo de Uso: Procesamiento de Logs con AWS

Supongamos que deseas analizar registros (logs) de aplicaciones web para obtener información valiosa. Puedes utilizar AWS para este escenario:

1. **Almacenamiento de Logs:** Utiliza Amazon S3 para almacenar los registros de aplicaciones web de manera escalable y duradera.
2. **Procesamiento con Amazon EMR:** Crea un clúster de Amazon EMR para procesar los registros utilizando Apache Spark. Puedes aplicar transformaciones y consultas para extraer información relevante.
3. **Almacenamiento de Resultados:** Guarda los resultados del procesamiento en Amazon Redshift para consultas posteriores y análisis enriquecidos.
4. **Visualización con Amazon QuickSight:** Utiliza Amazon QuickSight para crear paneles de control y visualizaciones interactivas que muestren información clave extraída de los logs.

1.5 AWS Academy

AWS Academy es un programa global de formación técnica respaldado por Amazon Web Services. Ofrece a estudiantes y educadores la capacitación y los recursos necesarios para desarrollar habilidades en la nube.

La misión de AWS Academy es proporcionar educación de alta calidad y actualizada en AWS, ayudando a los estudiantes a prepararse para carreras en la nube.

1.5.1 Beneficios

- Los estudiantes tienen la oportunidad de adquirir habilidades en la nube altamente demandadas por empleadores en una variedad de industrias. Ejemplo: Un estudio de empleo reciente mostró que el 80% de las empresas buscan profesionales de AWS.
- AWS Academy se enfoca en la práctica y proyectos reales, lo que permite a los estudiantes aplicar sus conocimientos en situaciones del mundo real.

1.5.2. Cursos Ofrecidos

- AWS Academy ofrece una amplia gama de cursos que cubren una variedad de temas, desde fundamentos de la nube hasta cursos especializados en áreas como desarrollo de aplicaciones y análisis de datos. Ejemplo: "Fundamentos de AWS Cloud" es un curso de nivel inicial que introduce los conceptos básicos de la nube.
- Muchos de los cursos de AWS Academy están diseñados para preparar a los estudiantes para las certificaciones de AWS, lo que les brinda una ventaja competitiva en el mercado laboral.

1.5.3. Learner Labs

- Los Learner Labs son espacios en línea donde los estudiantes pueden realizar prácticas relacionadas con AWS. Estos entornos ya están preconfigurados con recursos de AWS, como servidores virtuales y bases de datos, lo que permite a los estudiantes centrarse en aprender sin preocuparse por la infraestructura subyacente.

1.5.4. Certificaciones

Las certificaciones de AWS son reconocidas globalmente y validan las habilidades técnicas en la nube. Ejemplo: La certificación "AWS Certified Solutions Architect" demuestra la capacidad de diseñar sistemas escalables en AWS.